

# ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome

Michael J Taussig<sup>1</sup>, Oda Stoevesandt<sup>1</sup>, Carl A K Borrebaeck<sup>2</sup>, Andrew R Bradbury<sup>3</sup>, Dolores Cahill<sup>4</sup>, Christian Cambillau<sup>5</sup>, Antoine de Daruvar<sup>6</sup>, Stefan Dübel<sup>7</sup>, Jutta Eichler<sup>8</sup>, Ronald Frank<sup>8</sup>, Toby J Gibson<sup>9</sup>, David Gloriam<sup>10</sup>, Larry Gold<sup>11</sup>, Friedrich W Herberg<sup>12</sup>, Henning Hermjakob<sup>10</sup>, Jörg D Hoheisel<sup>13</sup>, Thomas O Joos<sup>14</sup>, Olli Kallioniemi<sup>15</sup>, Manfred Koegl<sup>16</sup>, Zoltán Konthur<sup>17</sup>, Bernhard Korn<sup>13</sup>, Elisabeth Kremmer<sup>17</sup>, Sylvia Krobitch<sup>18</sup>, Ulf Landegren<sup>19</sup>, Silvere van der Maarel<sup>20</sup>, John McCafferty<sup>21</sup>, Serge Muyldermans<sup>22</sup>, Per-Åke Nygren<sup>23</sup>, Sandrine Palcy<sup>6</sup>, Andreas Plückthun<sup>24</sup>, Bojan Polic<sup>25</sup>, Michael Przybylski<sup>26</sup>, Petri Saviranta<sup>15</sup>, Alan Sawyer<sup>27</sup>, David J Sherman<sup>28</sup>, Arne Skerra<sup>29</sup>, Markus Templin<sup>14</sup>, Marius Ueffing<sup>17</sup> & Mathias Uhlén<sup>23</sup>

ProteomeBinders is a new European consortium aiming to establish a comprehensive resource of well-characterized affinity reagents, including but not limited to antibodies, for analysis of the human proteome. Given the huge diversity of the proteome, the scale of the project is potentially immense but nevertheless feasible in the context of a pan-European or even worldwide coordination.

To explore the full complexity and function of the human proteome, it is essential to establish a comprehensive, characterized and standardized collection of specific binding molecules ('binders') directed against all individual human proteins, including variant forms and modifications. Primed with the knowledge of the human genome, such a systematic bank of affinity reagents would be a crucial precompetitive resource to understand and exploit the proteome<sup>1</sup>. Yet although affinity reagents are undeniably of central importance for proteomics, they presently cover only a very small fraction of the proteome, and even though there are many antibodies against some targets (for example, >900 antibodies against p53), there are none against the vast majority of proteins. Moreover, widely accepted standards for binder characterization are virtually nonexistent. Establishing a binder collection will not be an end in itself, but must be accompanied by development of high-throughput assay systems and new-generation protein-detection technologies. The benefits would include cost-effective reagent production and access as well as improved interlaboratory reproducibility,

and will have an impact on basic research and medicine as well as the biotechnology and pharmaceutical industries.

## ProteomeBinders: vision and goals

ProteomeBinders is a new European consortium with the vision of establishing an infrastructure resource of binding molecules for the entire human proteome, together with tools for their use and applications in studying proteome function and organization. When mature, the resource could be similar in nature to the American Type Culture Collection (ATCC) for cell lines, making the reagents available at cost and with no restrictions for research use. In this Commentary, we present the long-term goals of the ProteomeBinders initiative as well as the current activities of the consortium. The current 4-year, 1.8M€ (\$2.37M) initiative, funded by the European Commission 6<sup>th</sup> Framework Programme in the area of Research Infrastructures, is a 'Coordination Action' involving a network of 26 EU and 2 US partner institutions, leaders in the area of affinity-reagent production, characterization and application (see **Supplementary Table 1** online for a list of the lead participants in the ProteomeBinders consortium).

The consortium will coordinate several complementary activities:

1. Assessing the resources and methods required to develop a complete collection of binding molecules, from representation of the proteome in cDNA collections to binder selection and production.
2. Reviewing the properties of different molecular types of binders, including natural and recombinant antibodies, scaffold domains, peptides and nucleic acid aptamers, and linking them with proteomics tools and applications in research, diagnostics and therapeutics.
3. Establishing criteria and methods for universally applicable quality assessment and validation.
4. Defining standards for data representation and establishing a bioinformatics platform to display information on characterization of individual binders.
5. Planning the long-term production strategy and organization of the binder infrastructure.

The consortium will hold regular open workshops and disseminate information on the results of its activities through its website (<http://www.proteomebinders.org>), which will also contain a list of quality-assured binding reagents from all sources. In subsequent phases, the consortium aims to benchmark different types of binders and

Affiliations are listed at the end of the paper.  
e-mail: mike.taussig@bbsrc.ac.uk

production methods against defined sets of proteins to select those most appropriate for various applications. Embarking on the task of assembling the resource by systematically collecting and/or creating the tens of thousands of reagents needed will require an application for much more substantial funds, for example, from the next European Commission Framework Program, FP7, starting in 2007.

### Scale of the problem

The size of the human proteome (including splice variants, post-translational modifications, polymorphisms) is generally recognized to be at least an order of magnitude greater than the ~24,000 protein-coding genes (<http://www.ensembl.org>). This raises several central questions, presently being debated within the ProteomeBinders consortium: 'Is comprehensive coverage of the proteome realistic?' 'How should targets be prioritized—by biological or medical relevance, by unknown function, or according to other criteria (chromosome, etc)?' Another factor in the scale of the task is that several binders against each target will be required, depending on the nature of the samples (denatured or native), the biological status of the protein (post-translationally modified or not) or the detection mode of the assay (for example, sandwich configurations).

### Choices and challenges in developing a binder collection

The consortium will consider the following stages in binder production and application to reach a consensus for future action.

**Target generation.** Full-length proteins, probably the optimal targets for binder selection, can be expressed from cDNA collections<sup>2</sup> using a variety of systems (bacterial<sup>3</sup>, insect, mammalian, cell-free), but there is often limited success in obtaining them in soluble, correctly folded form<sup>4,5</sup>. Protein fragments or peptides are alternatives, especially combined with large-scale epitope prediction. 'Protein epitope signature tags' (PrESTs)<sup>6</sup> are genome-unique, nonrepetitive and nonhydrophobic protein subsequences, shown in the Swedish human proteome atlas project to be suitable for raising and affinity purifying polyclonal antibodies (<http://www.proteinatlas.org>)<sup>7</sup>. ProteomeBinders aims to integrate existing, but presently fragmented, bioinformatics tools into distributed and/or virtual facili-

ties that specifically address target-site selection within proteins. In the context of constructing a large-scale binder resource, an important goal will be to reduce the amount of target required (and hence also the cost), for example, by using microarrays for selection or using micro- or nanotechnology for specificity and affinity assays<sup>8</sup>.

**Molecular varieties of binders: antibodies and alternatives.** Antibodies are by far the most familiar and best understood affinity reagents—and generally the researcher's first choice—but not the only ones. ProteomeBinders unites expertise both on antibodies and alternative binding reagents with antibody-like specificity and affinity, including nonimmunoglobulin scaffolds<sup>9,10</sup> (Affibodies<sup>11</sup>, Anticalins<sup>12</sup>, designed Ankyrin repeat proteins<sup>13</sup> and others), nucleic acid aptamers<sup>14</sup>, peptides and chemical entities (**Box 1**). Antibodies (including monoclonals, monospecific polyclonals, camelid heavy chains<sup>15</sup>, and recombinant scFv fragments and single domains) are considerably more difficult to produce compared to high-yield bacterial expression of some alternative

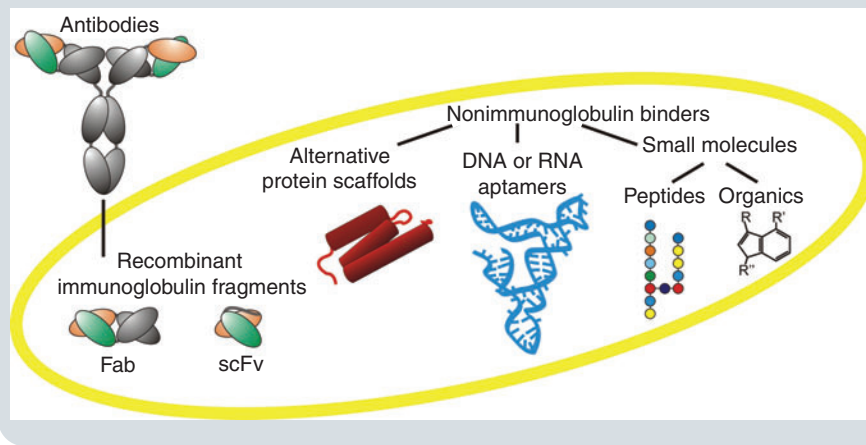
scaffolds<sup>13</sup>. The latter are also more robust and provide opportunities for engineering of functional properties. Among the attractions of antibodies are widespread competence in technologies, access to large libraries for recombinant selection<sup>16</sup> and availability of secondary reagents and detection systems. Recombinant binding molecules have the advantage of being completely described by their sequence, so that documentation and replication of experiments can be more objective. Though there may be a problem of wide acceptance of the alternatives to antibodies, users may well not be too concerned with the structure of the reagent, so long as it has been demonstrated to work in their particular application. Accordingly, the consortium will undertake benchmarking of the properties of alternative binders alongside conventional antibodies to define the 'right binder for the job'. Whatever the molecular species, sustainability will be a key factor: ultimately, a replenishable resource is required.

**Binder production methods and scaling.** The production of binding molecules for a systematic program can be contrasted in many

## BOX 1 DIFFERENT TYPES OF AFFINITY BINDERS

Antibodies and their fragments are the most familiar and widely used binding reagents. In comparison, protein scaffolds are often more robust (for example, in storage, reuse on columns) and give higher production yields in bacterial expression systems, partly due to the lack of cysteine residues. Aptamers stand out as being nucleic acid-based binding molecules, permitting simple production by synthesis or PCR, and distributable as sequence information. Libraries of small-molecule binders and peptides are accessible through combinatorial synthesis. Compared to a whole immunoglobulin, alternative binders are small, resulting in less steric hindrance, and are more readily used in intracellular applications. For the non-immunoglobulin binder formats, however, expertise is often limited to single laboratories and few libraries exist for distribution.

Inclusion into the yellow oval in the image indicates compatibility with *in vitro* selection methods (see **Box 2**).



**Table 1** | Comparison of binder generation for 'classical' hypothesis-driven research and systematic approaches (adapted from ref. 36)

	Hypothesis-driven	Systematic (proteomics)
Targets	Single protein	Large numbers of proteins
	Known and available (target production usually not part of the binder generation process)	Usually not known or available (to be provided in an integrated process)
	At least partially characterized	Properties unknown
Selection process	Number of steps is not an issue	Optimized for minimal number of steps
	Optimized for lowest failure rate	Several percent failure rate is not critical
Resulting binders	Individual assay conditions can be specified by end user	Standardization (for example, for parallel assays on arrays)
	Individual binders not selected or necessarily suited for parallel application under standardized conditions	Standardized selection and quality control schemes ensure suitability for systematic approaches (for example, minimized cross-reactivity)
	Large amounts of single binders	Small amounts of many binders (for example, for arrays)
Sources	Many companies	No commercial service available to handle the scale needed
Overall cost	Not a central issue	Optimized for minimal cost

respects with hypothesis-driven research (Table 1). For 'classical' antibodies, the production routes are either to raise polyclonals (purified for monospecificity) or hybridomas; throughput of monoclonals can be increased by immunization with antigen mixtures and selections on protein arrays<sup>17</sup>. For all molecular binder varieties, recombinant display library approaches can be applied and coupled with several possible selection methods. Systems with established track records such as phage display<sup>18</sup>, ribosome display<sup>19,20</sup>, cell-surface display, bacterial two-hybrid, functional colony screening, protein-fragment complementation<sup>21</sup>, SELEX<sup>22</sup> for aptamer selection and combinations of these methods<sup>23</sup> will be compared taking into account the molecular entity being selected and the intended downstream applications (Box 2). For example, if affinity is crucial, technologies with built-in evolution will be required (for example, ribosome display<sup>24</sup>), whereas if only 'some binder at some epitope' is needed, more technologies become available and selection is less stringent. For intracellular applications, binders should fold functionally in the reducing environment of the cell<sup>25</sup>. Other technology evaluation criteria include robustness, library creation and size, range of scaffolds that can be expressed, automation, and limits of scale and throughput. Contributions of the consortium will be to deliver effective protocols and actively check their robustness by rotating and annotating them between laboratories, and to identify potentially automatable steps, distinguishing those that are generic from the method-specific ones.

**Characterization and quality control.** The critical area of quality control is all too often sidelined. Validation will be a central issue;

there will be a requirement to demonstrate the quality of selection methods and binder formats, as well as of each individual binder. Although different binder types may have superior characteristics for defined applications, certain criteria (affinity, specificity<sup>8</sup> and cross-reactivity, native or denatured target, stability *in vivo* and *in vitro*, association and dissociation rate constants<sup>26,27</sup>) are applicable across all formats. The consortium will establish reference criteria for binder quality control and validation, which could eventually become a 'gold standard' in the research and commercial areas. Besides the classical tests of performance, for example, ELISA and western blotting, other validation methods range from protein and tissue microarrays to genomic correlations with transcript levels, gene knockouts, transgenesis and bioinformatic predictions (Table 2)<sup>7</sup>. High-throughput systems for initial specificity screens may be combined with advanced kinetic analyses for binder optimization<sup>28</sup>.

#### Linking binders to tools and applications.

The area of application is perhaps the most important criterion in choosing binder type, selection technology and characterization methods. Binder uses include high-throughput array methods (capture, tissue, lysate arrays) as well as the more classical techniques. For diagnostic and prognostic purposes, the pattern of information<sup>29</sup> that can be gained from target-binder interaction is more important than specificity of the binding event, as long as reproducibility and correlation with disease are high. In contrast, for functional analyses, where the global proteome-wide approach is particularly applicable, specificity is essential.

Accordingly, the parameters by which binders have to be evaluated can be very different. It is imperative to define applications beforehand and design or refine the selection process accordingly.

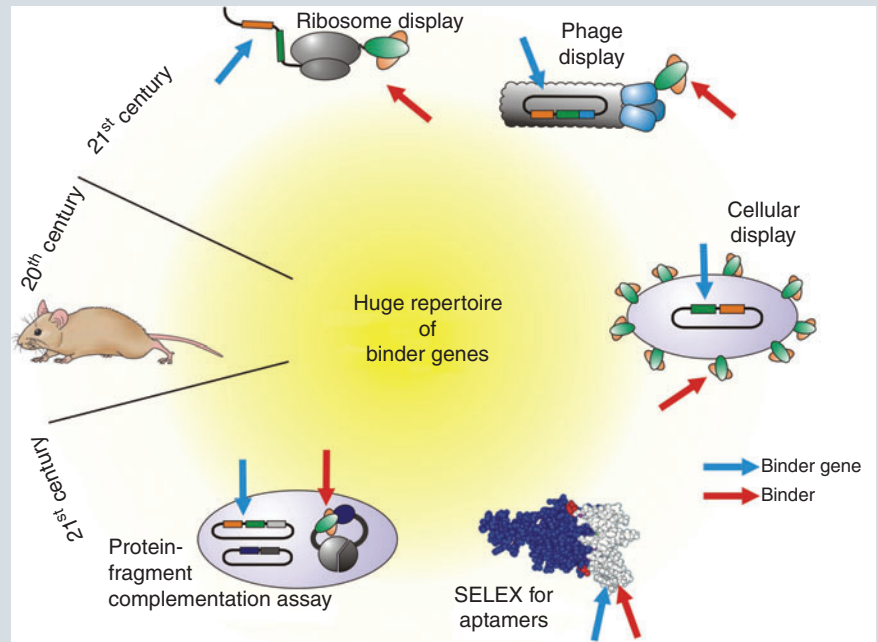
Novel methods to measure large sets of proteins using affinity reagents are becoming available, for example, fluorescence cross-correlation spectroscopy<sup>30</sup> and proximity ligation approaches, coupled with DNA amplification<sup>31–33</sup>. Further new-generation techniques must be evolved, particularly to improve sensitivity and specificity of detection (ultimately down to the single-molecule or single-cell level), the ability to perform highly multiplexed assays in individual samples, and to determine the spatial distribution of large numbers of molecules in cells and tissues.

**Bioinformatics resources.** ProteomeBinders will develop community standards for binder data representation in collaboration with the Human Proteome Organization (HUPO) Proteomics Standards Initiative<sup>34</sup> (<http://psidev.sourceforge.net/>). Crucial parameters of binder-target interactions have to be identified and an ontology of binder properties formally defined. These standards will be implemented in a comprehensive database of binders and other web resources, ideally in collaboration with other major binder providers. The database structure needs to anticipate the information to be captured, to allow retrieval of binders matching certain criteria and allow users a meaningful assessment of their performance. Basic searchable information should include the gene identifier of the target, the form of the target used for the binder generation or selection,

## BOX 2 21<sup>ST</sup> VERSUS 20<sup>TH</sup> CENTURY BINDER GENERATION TECHNOLOGIES: WHY USE *IN VITRO*-SELECTED AFFINITY REAGENTS?

*In vitro* recombinant selection systems link genotype and phenotype of the binding molecule. Features are:

- Scope of targets is not limited by the response of the immune system.
- Sequence information is sufficient to archive and recreate (and eventually distribute?) binders, which may avoid the need for a physical binder repository.
- Multiple mutagenesis-selection rounds: binders can be matured to picomolar affinity.
- Truly monoclonal, but can be made oligo- or polyclonal as desired.
- Functional domains and tags can be fused for downstream applications, for example: detection (GFP, alkaline phosphatase); multimerization to generate avidity and/or bispecificity (coiled coils); technology bridging (Fc fusions—access IgG-based technology, for example, secondary reagents); tethering (immobilization tags, for example, for array generation).
- Generation of fully human antibody fragments is possible, for example, for therapy.
- High throughput: standard selections yield  $\geq 10$  binders per target.
- Avoid use of animals.



a description of the molecular nature of the binder, the results from quality assurance and information about suggested applications and availability of the binder.

**Intellectual property.** Patent issues may well influence choices of binders and selec-

tion methods. Intellectual property rights to binders must be respected. It should be possible to find a model whereby the inventors of techniques for production of binders benefit from having their reagents selected by such a program. The relationship between an open access resource and commercial

activities will doubtless be a topic of future debate.

### Strategies for the long-term production phase

Clearly, new binders have to be made in very large numbers. For antibodies, outsourcing

**Table 2** | Validation criteria and methods for proteome binding molecules (adapted from ref. 7)

Approach based on	Examples of methods	Issues
Antigen (as used for immunization or <i>in vitro</i> selection)	ELISA, protein array, SPR	The antigen used in the selection is not always the target for later analyses.
Protein target (from natural sources, for example, cell lysate)	Western blot, IHC	Leaves margin of doubt in the absence of unpurified target as a control.
RNA (plausibility control to compare mRNA and protein expression levels)	Transcript profiling, <i>in situ</i> hybridization	Correlation of mRNA and protein levels is unknown.
Genetics (use genetic mutants or recombinant constructs to manipulate target in a defined way)	Transgenesis, RNAi, GFP fusions for comparison with IHC	Good validation if consistent with observed behavior of binder, for example, decreased binding upon RNAi knockdown of target.
DNA (use sequence information on target)	Bioinformatic analysis using predictive algorithms	Validation of localization in the absence or presence of transmembrane regions, localization signals and others
Epitope (compare results from two or more different binders to different areas, for example, PrESTs, of same target)	ELISA, protein array, western blot, IHC	Mutual validation of binders if corresponding binding behavior is detected.

IHC, immunohistochemistry; SPR, surface plasmon resonance.

to multiple sites can be considered. After adequate assessment, existing binders from the research community and commercial suppliers can also be incorporated. However, quality control and characterization should be retained within the consortium to ensure standardization. When trying to match the available funding with the task of generating more than 100,000 reagents, it becomes clear that new solutions and technical optimizations must be found. High-throughput approaches must be adopted at all levels, from protein expression and binder production to multiplexed assays and multiparameter tests<sup>35</sup>, and will be an integral part of any design of a binder-generation pipeline<sup>36</sup>.

### Networking in and beyond Europe

Other initiatives in the area of affinity reagents are the US National Cancer Institute proteome reagents program, focused mainly on cancer-related monoclonals<sup>37</sup> (<http://proteomics.cancer.gov>), the HUPRO antibody initiative (<http://www.hupo.org/research/hai/>) and the Swedish human proteome atlas project<sup>7</sup> (<http://www.proteinatlas.org>). Antibody Factory (<http://www.antibody-factory.de>), is a German national initiative to develop high-throughput recombinant antibody methods<sup>35,38</sup>. Additionally, small-molecule library initiatives have as their ultimate goal generating ligands for every human protein<sup>7</sup> function as anticipated by the US National Institutes of Health Molecular Libraries Screening Network<sup>39</sup> (<http://nihroadmap.nih.gov/molecularlibraries/>), Germany's ChemBioNet (<http://www.chembionet.de/>) and the new Spanish ChemBioBank (<http://www.pcb.ub.es/chembiobank/>). At the moment, these initiatives are independent and in some cases complementary; however, given the scale of the problem, it

would seem that in the future, the only way to tackle the task will be through coordination of activities.

Meanwhile, individual researchers have a critical role to play in shaping a resource that will eventually be theirs to use and they are expected to contribute to the effort by their comments and experience. The ProteomeBinders consortium welcomes comments and discussion from the wider community, via the website (<http://www.proteomebinders.org>) and participation in annual meetings (the next open workshop is in Alpbach, Austria, March 13–16 2007; details from [oda.stoevesandt@bbsrc.ac.uk](mailto:oda.stoevesandt@bbsrc.ac.uk)).

*Note: Supplementary information is available on the Nature Methods website.*

### ACKNOWLEDGMENTS

This article is dedicated to the memory of our consortium partner Jane Steel who tragically died last summer.

- Uhlen, M. & Ponten, F. *Mol. Cell. Proteomics* **4**, 384–393 (2005).
- Stelzl, U. *et al. Cell* **122**, 957–968 (2005).
- Dyson, M.R., Shadbolt, S.P., Vincent, K., Perera, R. & McCafferty, J. *BMC Biotechnol.* **4**, 32 (2004).
- Vincetelli, R., Canaan, S., Offant, J., Cambillau, C. & Bignon, C. *Anal. Biochem.* **346**, 77–84 (2005).
- Berrow, N.S. *et al. Acta Crystallogr. D Biol. Crystallogr.* **62**, 1218–1226 (2006).
- Larsson, K. *et al. J. Immunol. Methods* **315**, 110–120 (2006).
- Uhlen, M. *et al. Mol. Cell. Proteomics* **4**, 1920–1932 (2005).
- Poetz, O. *et al. Proteomics* **5**, 2402–2411 (2005).
- Binz, H.K., Amstutz, P. & Plückthun, A. *Nat. Biotechnol.* **23**, 1257–1268 (2005).
- Nygren, P.-A. & Skerra, A. *J. Immunol. Methods* **290**, 3–28 (2004).
- Wahlberg, E. *et al. Proc. Natl. Acad. Sci. USA* **100**, 3185–3190 (2003).
- Schlehuber, S. & Skerra, A. *Drug Discov. Today* **10**, 23–33 (2005).
- Binz, H.K. *et al. Nat. Biotechnol.* **22**, 575–582 (2004).
- Gold, L. in *The Aptamer Handbook: Functional Oligonucleotides and Their Applications* (ed., Klusman, S.) 461–468 (Wiley-VCH Verlag GmbH, Weinheim, 2006).
- De Genst, E. *et al. Proc. Natl. Acad. Sci. USA* **103**, 4586–4591 (2006).
- Knappik, A. *et al. J. Mol. Biol.* **296**, 57–86 (2000).
- De Masi, F. *et al. Proteomics* **5**, 4070–4081 (2005).
- Steiner, D., Forrer, P., Stumpp, M.T. & Plückthun, A. *Nat. Biotechnol.* **24**, 823–831 (2006).
- Hanes, J. & Plückthun, A. *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942 (1997).
- He, M. & Taussig, M.J. *Nucleic Acids Res.* **25**, 5132–5134 (1997).
- Pelletier, J.N., Arndt, K.M., Plückthun, A. & Michnick, S.W. *Nat. Biotechnol.* **17**, 683–690 (1999).
- Tuerk, C. & Gold, L. *Science* **249**, 505–510 (1990).
- Amstutz, P., Koch, H., Binz, H.K., Deuber, S.A. & Plückthun, A. *Protein Eng. Des. Sel.* **19**, 219–229 (2006).
- He, M., Cooley, N., Jackson, A. & Taussig, M.J. *Methods Mol. Biol.* **248**, 177–189 (2004).
- Lobato, M.N. & Rabbitts, T.H. *Curr. Mol. Med.* **4**, 519–528 (2004).
- Seitz, H. *et al. Proteomics* **6**, 5132–5139 (2006).
- Moll, D. *et al. J. Neural Transm.* **113**, 1015–1032 (2005).
- Kusnezow, W. *et al. Mol. Cell. Proteomics* **5**, 1681–1696 (2006).
- Borrebaeck, C.A.K. *Expert Opin. Biol. Ther.* **6**, 833–838 (2006).
- Stoevesandt, O., Köhler, K., Fischer, R., Johnston, I.C. & Brock, R. *Nat. Methods* **2**, 833–835 (2005).
- Gullberg, M. *et al. Proc. Natl. Acad. Sci. USA* **101**, 8420–8424 (2004).
- Fredriksson, S. *et al. Nat. Biotechnol.* **20**, 473–477 (2002).
- Soderberg, O. *et al. Nat. Methods* **3**, 995–1000 (2006).
- Orchard, S. *et al. Proteomics* **6**, 4439–4443 (2006).
- Konthur, Z., Hust, M. & Dübel, S. *Gene* **364**, 19–29 (2005).
- Hust, M. & Dübel, S. *Trends Biotechnol.* **22**, 8–14 (2004).
- Haab, B.B. *et al. Mol. Cell. Proteomics* **5**, 1996–2007 (2006).
- Walter, G., Konthur, Z. & Lehrach, H. *Comb. Chem. High Throughput Screen.* **4**, 193–205 (2001).
- Austin, C.P., Brady, L.S., Insel, T.R. & Collins, F.S. *Science* **306**, 1138–1139 (2004).

<sup>1</sup>Technology Research Group, The Babraham Institute, Cambridge CB22 3AT, UK. <sup>2</sup>Department of Immunotechnology, Lund University, SE-221 84 Lund, Sweden. <sup>3</sup>Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. <sup>4</sup>Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland. <sup>5</sup>Centre Nationale de la Recherche Scientifique, Universités Aix-Marseille I and II, 13288 Marseille Cedex 09, France. <sup>6</sup>Centre de Bioinformatique de Bordeaux, Université Bordeaux 2, 33076 Bordeaux Cedex, France. <sup>7</sup>Technical University Braunschweig, Institute of Biochemistry and Biotechnology, D-38106 Braunschweig, Germany. <sup>8</sup>Helmholtz Center for Infection Research, D-38124 Braunschweig, Germany. <sup>9</sup>European Molecular Biology Laboratory, D-69117 Heidelberg, Germany. <sup>10</sup>European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. <sup>11</sup>SomaLogic Inc., Boulder, Colorado 80301, USA. <sup>12</sup>Department of Biochemistry, University of Kassel, 34132 Kassel, Germany. <sup>13</sup>German Cancer Research Center, D-69120 Heidelberg, Germany. <sup>14</sup>Natural and Medical Sciences Institute at the University of Tübingen, 72770 Reutlingen, Germany. <sup>15</sup>VTT, Technical Research Center of Finland, FIN-20521 Turku, Finland. <sup>16</sup>Resource Center for Genome Research, D-69120 Heidelberg, Germany. <sup>17</sup>Institute of Human Genetics, GSF, National Research Center for Environment and Health, D-85764 Munich-Neuherberg, Germany. <sup>18</sup>Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. <sup>19</sup>Department of Genetics and Pathology, The Rudbeck Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden. <sup>20</sup>Leiden University Medical Center, 2300 RC Leiden, The Netherlands. <sup>21</sup>The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. <sup>22</sup>Vrije Universiteit Brussels, B-1050 Brussels, Belgium. <sup>23</sup>Royal Institute of Technology, AlbaNova University Center, SE-106 91 Stockholm, Sweden. <sup>24</sup>University of Zurich, Department of Biochemistry, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. <sup>25</sup>Medical Faculty University of Rijeka, Brace Branchetta 20, HR-51000 Rijeka, Croatia. <sup>26</sup>University of Konstanz, Department of Chemistry, D-78457 Konstanz, Germany. <sup>27</sup>European Molecular Biology Laboratory Monterotondo, Via Ramarini 32, Monterotondo-Scalo 00015, Italy. <sup>28</sup>Laboratoire Bordelais de Recherche en Informatique, 33405 Talence Cedex, France. <sup>29</sup>Lehrstuhl für Biologische Chemie, Technische Universität München, D-85350 Freising-Weißenstephan, Germany.